An Alternating Direction Method for Solving Convex Nonlinear Semidefinite Programming Problems[*]

Su Zhang[†] James Ang[‡] and Jie Sun[§]

April 14, 2011

Abstract

An alternating direction method is proposed for solving convex semidefinite optimization problems. The method only computes several metric projections at each iteration. Convergence analysis is presented and numerical experiments in solving matrix completion problems are reported.

Key Words. Alternating direction method, Convex semidefinite programming, Matrix completion, Variational inequalities

Dedication. This paper is dedicated to Professor Minyi Yue in celebration of his 90th birthday. Professor Yue is the pioneer in research of mathematical optimization in China. He trained an earliest group of mathematical programmers in China; Jie Sun was one of them.

[†]Department of Decision Sciences, National University of Singapore and Institute of Modern Management, Business School, Nankai University, PR China, Email:zhangsu@nankai.edu.cn.

[‡]Department of Decision Sciences, National University of Singapore, Email:bizangsk@nus.edu.sg.

[§]The corresponding author. Department of Decision Sciences and Risk Management Institute, National University of Singapore, Email:jsun@nus.edu.sg, Fax: +65-6779-2621.

1 Introduction

Optimization models play a very important role in operations research and management science. Optimization models with symmetric matrix variables are often referred to as semidefinite programs (SDPs). The study on these models has a relatively short history. Intensive studies on the theory, algorithms, and applications of semidefinite programs have only begun since the 1990s. However, so far most of the work has been concentrated on the linear case, where, except the semidefinite cone constraint, all other constraints as well as the objective function are linear with respect to the matrix variable.

A reasonable step forward, which is important both in theoretical development and in practical applications, is to develop methodologies for solving convex nonlinear semidefinite programs (CNLSDPs). This paper is concentrated on a prediction-correction alternating direction method for solving CNLSDP. The method is a first-order one. Therefore, the computational load is relatively light at each iteration. In fact, we will show that the main work of the algorithm can be reduced to computing metric projections on certain simple convex sets at each iteration. We will also report numerical results in applying this method for solving matrix completion problems.

Let $\mathbb{S}^n$ be the finite-dimensional Hilbert space of real symmetric matrices equipped with the Frobenius inner product. Let $\mathbb{S}^n_+$ ($\mathbb{S}^n_{++}$, respectively) be the subset of $\mathbb{S}^n$ consisting of all symmetric positive semidefinite (definite, respectively) matrices. Clearly, $\mathbb{S}^n_+$ is a convex cone and is therefore called the positive semidefinite cone. As a convention, we write $X \succeq 0$ ($X \succ 0$, respectively) to represent $X \in \mathbb{S}^n_+$ ($X \in \mathbb{S}^n_{++}$, respectively). We write $X \succeq Y$ or $Y \preceq X$ to represent $X - Y \succeq 0$. Similarly, we define $X \succ Y$ and $Y \prec X$.

We are concerned with solving the following CNLSDP

$$\min \; c_0(X) \quad \text{s.t. } X \succeq 0, \; c_i(X) \leq 0, \; i = 1, \ldots, m, \tag{1}$$

where $c_i : \mathbb{S}^n \rightarrow \Re, \; i = 0, 1, \ldots, m$, are convex continuously differentiable functions. Let $C_i(X), \; i = 0, 1, \ldots, m$, denote the first order derivative of $c_i(X)$. Furthermore, we require the operator $C_0(\cdot)$ to be Lipschitz continuous with a constant L.

Recently, some methods have been proposed for solving nonlinear semidefinite programs (NLSDPs). Kocvara and Stingl [24] developed a code (PENNON), where the augmented Lagrangian method is used. Sun et al. [35] analyzed the convergence rate for the augmented Lagrangian method in the general nonlinear SDP setting. A smoothing Newton method for NLSDPs, which is a second-order algorithm, is considered in Sun et al. [36]. A variant of the smoothing Newton methods is subsequently studied in [25]. An analytic center cutting plane method is investigated by Sun et al. [37, 40], which can in principle be used for solving CNLSDPs of certain type. In Jarre [21], Leibfitz and Mostafa [28], and Yamashita et al. [43], interior methods are discussed. In addition, Gowda and his collaborators have extensively studied complementarity problems in the general symmetric cone setting, e.g., [17, 18], which are closely related to the solution of NLSDPs.

Among the first-order approaches, a successive linearization method was considered by Fares, Noll, and Apkarian [11], by Correa and Ramirez [6], and by Kanzow et al. [22]. Noll and Apkarian [31, 32] also suggested spectral bundle methods. Our proposed algorithm is different from the above methods in two aspects. Firstly, we aim at convex nonlinear problems, hence the convergence result is stronger than the general nonlinear case. Secondly, as the proposed

approach is a projection based algorithm it does not require to solve SDP problems at each iteration. As such, it is simple to implement and is efficient, compared to the current first-order methods.

The paper is organized as follows. We present a brief review of the alternating direction method and propose a prediction-correction version of it for solving CNLSDP in Section 2. In Section 3 we present the convergence proof. Section 4 includes our preliminary computational results. Finally, we make some concluding remarks in Section 5.

2 The Algorithm

2.1. Review of the Alternating Direction Method

By introducing

$$Y_i = X \text{ and } \Omega_i = \{Y_i : c_i(Y_i) \leq 0\}, \ i = 1, \ldots, m, \tag{2}$$

we rewrite (1) equivalently as

$$\begin{aligned} \min \quad & c_0(X) \\ \text{s.t.} \quad & X = Y_i, \ Y_i \in \Omega_i, \ i = 1, \ldots, m \\ & X \succeq 0 \end{aligned} \tag{3}$$

The Lagrange dual of problem (3) is

$$\max_{\lambda_i} \min_{X \succeq 0, Y_i \in \Omega_i} L(X, Y_1, \ldots, Y_m, \lambda_1, \ldots, \lambda_m) \equiv c_0(X) - \sum_{i=1}^{m} \langle \lambda_i, X - Y_i \rangle. \tag{4}$$

Notice that the Lagrange multipliers λ_i are symmetric matrices. It is well known that under mild constraint qualifications (e.g. Slater's condition), strong duality holds and hence, X^* is a solution of (3) if and only if there exist $\lambda_i^* \in \mathbb{S}^n$ such that $(X^*, Y_i^*, \lambda_i^*)$ satisfies

$$\begin{cases} \left\langle X - X^*, C_0(X^*) - \sum_{i=1}^{m} \lambda_i^* \right\rangle \geq 0, \ \forall \ X \in \mathbb{S}_+^n \\ \langle Y_i - Y_i^*, \lambda_i^* \rangle \geq 0, \ \forall \ Y_i \in \Omega_i, \ i = 1, \ldots, m \\ X^* = Y_i^*, \ i = 1, \ldots, m \end{cases} \tag{5}$$

Problem (5) is a variational inequality problem with a special structure. The variables (X, Y_i, λ_i) are symmetric matrices, and the underlying set $\mathbb{S}_+^n$ is convex and non-polyhedral. We assume that the metric projections on Ω_is can be readily computed. Let $\Pi_S(\cdot)$ stand for the metric projection onto a set S.

For convenience, we state the basic assumption to guarantee that problem (3) under consideration is solvable.

Assumption 2.1. *The solution set $(X^*, Y_i^*, \lambda_i^*)$ of KKT system (5) of problem (3) is nonempty.*

Among the first-order approaches for solving large optimization problems, the augmented Lagrangian method has desirable convergence properties [35]. However, a quadratic penalty term is added to the Lagrangian function (4). This additional term is usually not separable respective to X and Y_i, which makes the augmented Lagrangian method more difficult to implement. To overcome this difficulty, the alternating direction method is introduced. The alternating direction method generally consists of three steps.

(I) Minimize the augmented Lagrangian function with respective to X only.

(II) Minimize the augmented Lagrangian function with respective to Y_i only.

(III) Update the Lagrangian multipliers λ_i.

Repeat (I), (II), and (III) until a stopping criterion is satisfied.

The alternating direction method can be seen as the block Gauss-Seidel variant of the augmented Lagrangian approach. The fundamental principle involved is to use the most recent information as it is available. Furthermore, it is very suitable for parallel computation in a data parallel environment. The alternating direction method was probably first considered by Gabay [15] and Gabay and Mercier [16]. As shown in [29], the alternating direction method is actually an instance of the Douglas-Rachford splitting procedure of monotone operators [7]. It is also related to the progressive hedging algorithm of Rockafellar and Wets [34]. The alternating direction method has been studied quite extensively in the settings of optimization and numerical analysis. Eckstein [8] and Kontogiorgis [26] gave a detailed analysis of alternating direction methods and tested their efficiency using numerical experiments in the parallel computation environment. Some versions of the alternating direction methods for solving different convex optimization problems appeared in [9, 14, 27]. Further studies of the alternating direction method can be found, for instance, in [19, 20]. He et al. [19] proposed an inexact alternating direction method with flexible conditions for structured monotone variational inequalities. Recently, He et al. [20] considered alternating projection-based prediction-correction methods for structured variational inequalities. All of the work above, however, was devoted to vector optimization problems. Independent from our work, Wen et al. [42] applied the idea in the works of Burer and Vandenbussche [3] and Povh et al. [33] and proposed an alternating direction method for linear semidefinite optimization problems. It appears to be new to apply the idea of alternating direction method to develop methods for solving CNLSDP problems.

When applied to problem (3), the alternating direction method becomes the following.

Algorithm 2.2. The Alternating Direction Method for CNLSDP

Do at each iteration until a stopping criterion is met

Step 1. $\left(X^k, Y_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, Y_i^k, \lambda_i^k\right)$, *where*

$$\left\langle X - \widetilde{X}^k, C_0\left(\widetilde{X}^k\right) - \sum_{i=1}^{m}\left(\lambda_i^k - \beta_i\left(\widetilde{X}^k - Y_i^k\right)\right)\right\rangle \geq 0, \ \forall X \succeq 0. \tag{6}$$

$\beta_i, \ i = 1, \ldots, m$, *is a certain positive scalar.*

Step 2. $\left(\widetilde{X}^k, Y_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, \widetilde{Y}_i^k, \lambda_i^k\right)$, $i = 1, \ldots, m$, *where*

$$\left\langle Y_i - \widetilde{Y}_i^k, \lambda_i^k - \beta_i\left(\widetilde{X}^k - \widetilde{Y}_i^k\right)\right\rangle \geq 0, \ \forall Y_i \in \Omega_i. \tag{7}$$

Step 3. $\left(\widetilde{X}^k, \widetilde{Y}_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, \widetilde{Y}_i^k, \widetilde{\lambda}_i^k\right)$, $i = 1, \ldots, m$, *where*

$$\widetilde{\lambda}_i^k = \lambda_i^k - \beta_i\left(\widetilde{X}^k - \widetilde{Y}_i^k\right), \tag{8}$$

and update $(X^k, Y_i^k, \lambda_i^k)$ *by* $(\widetilde{X}^k, \widetilde{Y}_i^k, \widetilde{\lambda}_i^k)$.

2.2. The Prediction-Correction Alternating Direction Method for CNLSDP

If we implement Algorithm 2.2 for solving CNLSDP problems, we would have to solve sub-variational inequality problems on matrix spaces at each iteration. Although there are a number of methods for solving monotone variational inequalities, in many occasions it is not an easy task. As a matter of fact, there seems to be little justification for the effort of obtaining the solutions of these sub-problems at each iteration. Therefore, we modify the original alternating direction method to make the implementation of each iteration much easier. Specifically, after the modifications the main computational load of each iteration is only the metric projections onto convex sets in the matrix space.

In the following, we will convert Step 1 and Step 2 to simple projection operations. For this purpose, all we need is the following fact from convex geometry.

Lemma 2.3. *([23] Theorem 2.3) Let Ω be a closed convex set in a Hilbert space. Then*

$$\langle z - y, y - x \rangle \geq 0, \quad \forall z \in \Omega \iff y \in P_\Omega(x). \tag{9}$$

By using this lemma, we can see that (7) is equivalent to the following nonlinear equation

$$\widetilde{Y}_i^k = \Pi_{\Omega_i} \left[\widetilde{Y}_i^k - \alpha_i \left(\lambda_i^k - \beta_i \left(\widetilde{X}^k - \widetilde{Y}_i^k \right) \right) \right],$$

where $\Pi_\Omega(U)$ stands for the projection of U onto Ω and α_i can be any positive number. Thus by choosing $\alpha_i = \frac{1}{\beta_i}$, the right hand side $\widetilde{Y}_i^k$s are cancelled. That is, in order to solve (7), we only have to compute

$$\widetilde{Y}_i^k = \Pi_{\Omega_i} \left[\widetilde{X}^k - \frac{1}{\beta_i} \lambda_i^k \right]. \tag{10}$$

However, it does not work for (6) since it is generally impossible to select an α so that the right hand side $\widetilde{X}^k$s are cancelled in

$$\widetilde{X}^k = \Pi_{\mathbb{S}_+^n} \left[\widetilde{X}^k - \alpha \left(C_0 \left(\widetilde{X}^k \right) - \sum_{i=1}^m \left(\lambda_i^k - \beta_i \left(\widetilde{X}^k - Y_i^k \right) \right) \right) \right].$$

We therefore suggest the following approximate approach. Let

$$R \left(X^k, \widetilde{X}^k \right) = \left(1 - \alpha \sum_{i=1}^m \beta_i \right) \left(X^k - \widetilde{X}^k \right) - \alpha \left(C_0 \left(X^k \right) - C_0 \left(\widetilde{X}^k \right) \right),$$

where we choose positive scalar α so that

$$\alpha \leq \frac{\eta}{L + \eta \sum_{i=1}^m \beta_i} \tag{11}$$

with a certain $0 < \eta < 1$ (L is the Lipschitz constant of $C_0(\cdot)$). Note that

$$\Pi_{\mathbb{S}_+^n} \left[\widetilde{X}^k - \alpha \left(C_0 \left(\widetilde{X}^k \right) - \sum_{i=1}^m \left(\lambda_i^k - \beta_i \left(\widetilde{X}^k - Y_i^k \right) \right) \right) + R \left(X^k, \widetilde{X}^k \right) \right]$$

$$= \Pi_{\mathbb{S}_+^n} \left[X^k - \alpha \left(C_0 \left(X^k \right) - \sum_{i=1}^m \left(\lambda_i^k - \beta_i \left(X^k - Y_i^k \right) \right) \right) \right]. \tag{12}$$

Now all implicit parts within the projections have been successfully cancelled. However, we cannot prove the convergence by just letting $X^{k+1} = (12)$. Instead we use it as the predictor and will correct it in the correction phase. In summary, the prediction-correction alternating direction method is thus given as follows.

Algorithm 2.4. The Prediction-Correction Alternating Direction Method for CNLSDP

Do at each iteration until a stopping criterion is met

The Prediction Phase:

Step 1. $\left(X^k, Y_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, Y_i^k, \lambda_i^k\right)$, *where*

$$\widetilde{X}^k = \Pi_{\mathbb{S}_+^n}\left[X^k - \alpha\left(C_0\left(X^k\right) - \sum_{i=1}^m\left(\lambda_i^k - \beta_i\left(X^k - Y_i^k\right)\right)\right)\right], \tag{13}$$

where α and β_i, $i = 1, \ldots, m$, are certain positive scalars satisfying (11).

Step 2. $\left(\widetilde{X}^k, Y_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, \widetilde{Y}_i^k, \lambda_i^k\right)$, $i = 1, \ldots, m$, *where*

$$\widetilde{Y}_i^k = \Pi_{\Omega_i}\left[\widetilde{X}^k - \frac{1}{\beta_i}\lambda_i^k\right]. \tag{14}$$

Step 3. $\left(\widetilde{X}^k, \widetilde{Y}_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, \widetilde{Y}_i^k, \widetilde{\lambda}_i^k\right)$, $i = 1, \ldots, m$, *where*

$$\widetilde{\lambda}_i^k = \lambda_i^k - \beta_i\left(\widetilde{X}^k - \widetilde{Y}_i^k\right). \tag{15}$$

The Correction Phase:

Step 4. $\left(\widetilde{X}^k, \widetilde{Y}_i^k, \widetilde{\lambda}_i^k\right) \to \left(X^{k+1}, Y_i^{k+1}, \lambda_i^{k+1}\right)$, *where*

$$
\begin{aligned}
X^{k+1} &= \Pi_{\mathbb{S}_+^n}\left[X^k - \gamma^k R\left(X^k, \widetilde{X}^k\right)\right], & (16)\\
Y_i^{k+1} &= \Pi_{\Omega_i}\left[Y_i^k - \gamma^k\left(Y_i^k - \widetilde{Y}_i^k\right)\right], & i = 1, \ldots, m, & (17)\\
\lambda_i^{k+1} &= \lambda_i^k - \gamma^k\left(\lambda_i^k - \widetilde{\lambda}_i^k\right), & i = 1, \ldots, m. & (18)
\end{aligned}
$$

The positive scalar γ^k is a step-length and its choice will be given later.

3 Convergence Analysis

Theorem 3.1. *The sequence $\left\{X^k, \widetilde{X}^k, Y_i^k, \widetilde{Y}_i^k, \lambda_i^k, \widetilde{\lambda}_i^k\right\}$ generated by the prediction-correction alternating direction method for CNLSDP satisfies*

$$\sum_{i=1}^m \frac{\alpha}{\beta_i}\left\langle \widetilde{\lambda}_i^k - \lambda_i^*, \lambda_i^k - \widetilde{\lambda}_i^k\right\rangle + \alpha\sum_{i=1}^m \beta_i\left\langle \widetilde{Y}_i^k - Y_i^*, Y_i^k - \widetilde{Y}_i^k\right\rangle + \left\langle \widetilde{X}^k - X^*, R\left(X^k, \widetilde{X}^k\right)\right\rangle \geq 0, \tag{19}$$

where $(X^*, Y_i^*, \lambda_i^*)$ are defined as in (5).

5

Proof. Using (5) and (7), we have

$$\left\langle \widetilde{Y}_i^k - Y_i^*, \lambda_i^* - \widetilde{\lambda}_i^k \right\rangle \geq 0 \tag{20}$$

Similarly, we get

$$\left\langle \widetilde{Y}_i^k - Y_i^k, \lambda_i^k - \widetilde{\lambda}_i^k \right\rangle \geq 0 \tag{21}$$

Note that (12) can be written equivalently as

$$\left\langle X - \widetilde{X}^k, \alpha \left(C_0\left(\widetilde{X}^k\right) - \sum_{i=1}^m \left(\lambda_i^k - \beta_i\left(\widetilde{X}^k - Y_i^k\right)\right)\right) - R\left(X^k, \widetilde{X}^k\right) \right\rangle \geq 0, \ \forall X \in \mathbb{S}_+^n$$

in view of the relationship (9). Setting $X = X^*$ in it and using (15), we obtain

$$\left\langle \widetilde{X}^k - X^*, -\alpha \left(C_0\left(\widetilde{X}^k\right) - \sum_{i=1}^m \widetilde{\lambda}_i^k \right) - \alpha \sum_{i=1}^m \beta_i\left(\widetilde{Y}_i^k - Y_i^k\right) + R\left(X^k, \widetilde{X}^k\right) \right\rangle \geq 0. \tag{22}$$

Let $X = \widetilde{X}^k$ in inequality (5). Then

$$\alpha \left\langle \widetilde{X}^k - X^*, C_0\left(X^*\right) - \sum_{i=1}^m \lambda_i^* \right\rangle \geq 0 \tag{23}$$

Adding (22) and (23) together, it follows that

$$\alpha \left\langle \sum_{i=1}^m \left(\widetilde{\lambda}_i^k - \lambda_i^*\right), \widetilde{X}^k - X^* \right\rangle + \alpha \left\langle \sum_{i=1}^m \beta_i\left(Y_i^k - \widetilde{Y}_i^k\right), \widetilde{X}^k - X^* \right\rangle$$
$$+ \ \left\langle \widetilde{X}^k - X^*, R\left(X^k, \widetilde{X}^k\right) \right\rangle \geq \alpha \left\langle \widetilde{X}^k - X^*, C_0\left(\widetilde{X}^k\right) - C_0\left(X^*\right) \right\rangle \geq 0. \tag{24}$$

It follows from (20), (21) and (24) that

$$\alpha \left\langle \sum_{i=1}^m \left(\widetilde{\lambda}_i^k - \lambda_i^*\right), \widetilde{X}^k - X^* \right\rangle + \alpha \left\langle \sum_{i=1}^m \beta_i\left(Y_i^k - \widetilde{Y}_i^k\right), \widetilde{X}^k - X^* \right\rangle$$
$$+ \left\langle \widetilde{X}^k - X^*, R\left(X^k, \widetilde{X}^k\right) \right\rangle + \alpha \sum_{i=1}^m \left\langle \widetilde{Y}_i^k - Y_i^*, \lambda_i^* - \widetilde{\lambda}_i^k \right\rangle$$
$$+ \alpha \sum_{i=1}^m \left\langle \widetilde{Y}_i^k - Y_i^k, \lambda_i^k - \widetilde{\lambda}_i^k \right\rangle$$
$$= \sum_{i=1}^m \frac{\alpha}{\beta_i} \left\langle \widetilde{\lambda}_i^k - \lambda_i^*, \lambda_i^k - \widetilde{\lambda}_i^k \right\rangle + \alpha \sum_{i=1}^m \beta_i \left\langle \widetilde{Y}_i^k - Y_i^*, Y_i^k - \widetilde{Y}_i^k \right\rangle$$
$$+ \left\langle \widetilde{X}^k - X^*, R\left(X^k, \widetilde{X}^k\right) \right\rangle \geq 0.$$

The proof is completed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \Box$

The following is the main theorem.

Theorem 3.2. *The sequence $\left\{X^k\right\}$ generated by the prediction-correction alternating direction method for CNLSDP converges to a solution point X^*.*

Proof. We denote

$$W \equiv \begin{pmatrix} X \\ Y_i \\ \lambda_i \end{pmatrix}; \quad G \equiv \begin{pmatrix} I & 0 & 0 \\ 0 & \alpha H & 0 \\ 0 & 0 & \alpha H^{-1} \end{pmatrix}$$

where I is the identical matrix and H is a diagonal matrix with β_i on its diagonal. Clearly, G is positive definite. We define the G-inner product of W and W' as

$$\langle W, W' \rangle_G \equiv \alpha \sum_{i=1}^{m} \frac{1}{\beta_i} \langle \lambda_i, \lambda_i' \rangle + \alpha \sum_{i=1}^{m} \beta_i \langle Y_i, Y_i' \rangle + \langle X, X' \rangle$$

and the associated G-norm as

$$\|W\|_G \equiv \left(\alpha \sum_{i=1}^{m} \frac{1}{\beta_i} \|\lambda_i\|^2 + \alpha \sum_{i=1}^{m} \beta_i \|Y_i\|^2 + \|X\|^2 \right)^{\frac{1}{2}}.$$

Observe that solving the optimal condition (5) for problem (3) is equivalent to finding a zero point of the residual function

$$\|e(W)\|_G \equiv \left\| \begin{array}{c} X - \Pi_{\mathbb{S}_+^n} \left[X - \alpha \left(C_0(X) - \sum_{i=1}^{m} \lambda_i \right) \right] \\ Y_i - \Pi_{\Omega_i} \left[Y_i - \alpha_i \lambda_i \right] \\ \beta_i (X - Y_i) \end{array} \right\|_G .$$

Then we have

$$\widetilde{X}^k = \Pi_{\mathbb{S}_+^n} \left[\widetilde{X}^k - \alpha \left(C_0\left(\widetilde{X}^k \right) - \sum_{i=1}^{m} \left(\lambda_i^k - \beta_i \left(\widetilde{X}^k - Y_i^k \right) \right) \right) + R\left(X^k, \widetilde{X}^k \right) \right]$$

$$\Rightarrow \left\| e\left(\widetilde{W}^k \right) \right\|_G \leq \left\| \begin{array}{c} -\alpha \sum_{i=1}^{m} \beta_i \left(\widetilde{Y}_i^k - Y_i^k \right) + R\left(X^k, \widetilde{X}^k \right) \\ 0 \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{array} \right\|_G$$

(non-expansion of projection)

$$\leq \left\| \begin{array}{c} -\alpha \sum_{i=1}^{m} \beta_i \left(\widetilde{Y}_i^k - Y_i^k \right) + \left(1 - \alpha \sum_{i=1}^{m} \beta_i \right) \left(X^k - \widetilde{X}^k \right) - \alpha \left(C_0\left(X^k \right) - C_0\left(\widetilde{X}^k \right) \right) \\ 0 \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{array} \right\|_G$$

$$\leq a \left\| W^k - \widetilde{W}^k \right\|_G, \tag{25}$$

(because of (11))

where a is a positive constant. From (11), there also holds the following inequality

$$\left\langle \begin{pmatrix} X^k - \widetilde{X}^k \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix}, \begin{pmatrix} R\left(X^k, \widetilde{X}^k\right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix} \right\rangle_G$$

$$= \left\langle X^k - \widetilde{X}^k, \left(1 - \alpha \sum_{i=1}^m \beta_i\right)\left(X^k - \widetilde{X}^k\right) - \alpha\left(C_0\left(X^k\right) - C_0\left(\widetilde{X}^k\right)\right)\right\rangle$$

$$+\alpha \sum_{i=1}^m \beta_i \left\langle Y_i^k - \widetilde{Y}_i^k, Y_i^k - \widetilde{Y}_i^k \right\rangle + \alpha \sum_{i=1}^m \frac{1}{\beta_i} \left\langle \lambda_i^k - \widetilde{\lambda}_i^k, \lambda_i^k - \widetilde{\lambda}_i^k \right\rangle$$

$$\geq \left(1 - \alpha \sum_{i=1}^m \beta_i\right)(1 - \eta)\left\|X^k - \widetilde{X}^k\right\|^2 + \alpha \sum_{i=1}^m \beta_i \left\|Y_i^k - \widetilde{Y}_i^k\right\|^2 + \alpha \sum_{i=1}^m \frac{1}{\beta_i}\left\|\lambda_i^k - \widetilde{\lambda}_i^k\right\|^2$$

$$\geq \left(1 - \alpha \sum_{i=1}^m \beta_i\right)(1 - \eta)\left\|\widetilde{W}^k - W^k\right\|_G^2. \tag{26}$$

Thus,

$$
\begin{aligned}
\left\| W^{k+1} - W^* \right\|_G^2 &= \left\| X^{k+1} - X^* \right\|^2 + \alpha \sum_{i=1}^m \beta_i \left\| Y_i^{k+1} - Y_i^* \right\|^2 + \alpha \sum_{i=1}^m \frac{1}{\beta_i} \left\| \lambda_i^{k+1} - \lambda_i^* \right\|^2 \\
&\leq \left\| X^k - \gamma^k R\left(X^k, \widetilde{X}^k \right) - X^* \right\|^2 + \alpha \sum_{i=1}^m \beta_i \left\| Y_i^k - \gamma^k \left(Y_i^k - \widetilde{Y}_i^k \right) - Y^* \right\|^2 \\
&\quad + \alpha \sum_{i=1}^m \frac{1}{\beta_i} \left\| \lambda_i^k - \gamma^k \left(\lambda_i^k - \widetilde{\lambda}_i^k \right) - \lambda_i^* \right\|^2 \\
&= \left\| X^k - X^* \right\|^2 - 2\gamma^k \left\langle X^k - X^*, R\left(X^k, \widetilde{X}^k \right) \right\rangle + \left(\gamma^k \right)^2 \left\| R\left(X^k, \widetilde{X}^k \right) \right\|^2 \\
&\quad + \alpha \sum_{i=1}^m \beta_i \left\{ \left\| Y_i^k - Y_i^* \right\|^2 - 2\gamma^k \left\langle Y_i^k - Y_i^*, Y_i^k - \widetilde{Y}_i^k \right\rangle + \left(\gamma^k \right)^2 \left\| Y_i^k - \widetilde{Y}_i^k \right\|^2 \right\} \\
&\quad + \alpha \sum_{i=1}^m \frac{1}{\beta_i} \left\{ \left\| \lambda_i^k - \lambda_i^* \right\|^2 - 2\gamma^k \left\langle \lambda_i^k - \lambda_i^*, \lambda_i^k - \widetilde{\lambda}_i^k \right\rangle + \left(\gamma^k \right)^2 \left\| \lambda_i^k - \widetilde{\lambda}_i^k \right\|^2 \right\} \\
&\leq \left\| X^k - X^* \right\|^2 - 2\gamma^k \left\langle X^k - \widetilde{X}^k, R\left(X^k, \widetilde{X}^k \right) \right\rangle + \left(\gamma^k \right)^2 \left\| R\left(X^k, \widetilde{X}^k \right) \right\|^2 \\
&\quad + \alpha \sum_{i=1}^m \beta_i \left\{ \left\| Y_i^k - Y_i^* \right\|^2 - 2\gamma^k \left\langle Y_i^k - \widetilde{Y}_i^k, Y_i^k - \widetilde{Y}_i^k \right\rangle + \left(\gamma^k \right)^2 \left\| Y_i^k - \widetilde{Y}_i^k \right\|^2 \right\} \\
&\quad + \alpha \sum_{i=1}^m \frac{1}{\beta_i} \left\{ \left\| \lambda_i^k - \lambda_i^* \right\|^2 - 2\gamma^k \left\langle \lambda_i^k - \widetilde{\lambda}_i^k, \lambda_i^k - \widetilde{\lambda}_i^k \right\rangle + \left(\gamma^k \right)^2 \left\| \lambda_i^k - \widetilde{\lambda}_i^k \right\|^2 \right\} \\
&= \left\| W^k - W^* \right\|_G^2 - 2\gamma^k \left\langle \begin{pmatrix} X^k - \widetilde{X}^k \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix}, \begin{pmatrix} R\left(X^k, \widetilde{X}^k \right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix} \right\rangle_G \\
&\quad + \left(\gamma^k \right)^2 \left\| \begin{pmatrix} R\left(X^k, \widetilde{X}^k \right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix} \right\|_G^2 \qquad (27)
\end{aligned}
$$

$$
= \left\| W^k - W^* \right\|_G^2 - \nu(2-\nu)\gamma_*^k \left\langle \begin{pmatrix} X^k - \widetilde{X}^k \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix}, \begin{pmatrix} R\left(X^k, \widetilde{X}^k \right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix} \right\rangle_G
$$

$$
\left(\text{where } \gamma^k \equiv \nu\gamma_*^k = \nu \frac{\left\langle \begin{pmatrix} X^k - \widetilde{X}^k \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix}, \begin{pmatrix} R\left(X^k, \widetilde{X}^k \right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix} \right\rangle_G}{\left\| \begin{pmatrix} R\left(X^k, \widetilde{X}^k \right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k \end{pmatrix} \right\|_G^2} \text{ and } 0 < \nu < 2 \right)
$$

$$
\leq \left\| W^k - W^* \right\|_G^2 - \nu\left(2 - \nu \right)\gamma_*^k \left(1 - \alpha \sum_{i=1}^m \beta_i \right) (1-\eta) \left\| \widetilde{W}^k - W^k \right\|_G^2 \qquad (28)
$$

$$
\leq \left\| W^k - W^* \right\|_G^2 - \nu\left(2 - \nu \right)\gamma_*^k \left(1 - \alpha \sum_{i=1}^m \beta_i \right) (1-\eta) \frac{1}{a^2} \left\| e\left(\widetilde{W}^k \right) \right\|_G^2, \qquad (29)
$$

where (26) was used to deduce (28) and (25) was used to obtain (29), respectively. From the inequality above, we have

$$\left\| W^{k+1} - W^* \right\|_G^2 \leq \left\| W^k - W^* \right\|_G^2 \leq \ldots \leq \left\| W^0 - W^* \right\|_G^2. \tag{30}$$

That is, the sequence $\{W^k\}$ is bounded. It follows from (28) that

$$\sum_{k=0}^{\infty} \nu (2 - \nu) \gamma_*^k \left(1 - \alpha \sum_{i=1}^{m} \beta_i \right) (1 - \eta) \left\| \widetilde{W}^k - W^k \right\|_G^2 < +\infty.$$

This implies that

$$\lim_{k \to \infty} \left\| \widetilde{W}^k - W^k \right\|_G = 0.$$

Thus the sequence $\left\{ \widetilde{W}^k \right\}$ is also bounded. Then there exists at least one cluster point of $\left\{ \widetilde{W}^k \right\}$.

It also follows from (29) that

$$\sum_{k=0}^{\infty} \nu (2 - \nu) \gamma_*^k \left(1 - \alpha \sum_{i=1}^{m} \beta_i \right) (1 - \eta) \frac{1}{a^2} \left\| e \left(\widetilde{W}^k \right) \right\|_G^2 < +\infty.$$

This implies that

$$\lim_{k \to \infty} \left\| e \left(\widetilde{W}^k \right) \right\|_G = 0.$$

Let $\overline{W}$ be a cluster point of $\left\{ \widetilde{W}^k \right\}$, and let $\left\{ \widetilde{W}^{k_j} \right\}$ be a corresponding subsequence converging to $\overline{W}$. Therefore,

$$\lim_{j \to \infty} \left\| \overline{W} - W^{k_j} \right\|_G = \lim_{j \to \infty} \left\| \widetilde{W}^{k_j} - W^{k_j} \right\|_G = 0.$$

Also note that

$$\left\| e \left(\overline{W} \right) \right\|_G = \lim_{j \to \infty} \left\| e \left(\widetilde{W}^{k_j} \right) \right\|_G = 0,$$

which means that $\overline{W}$ is a zero point of the residual function. Therefore $\overline{W}$ satisfies (5). Setting $W^* = \overline{W}$ in (30), we have

$$\left\| W^{k+1} - \overline{W} \right\|_G^2 \leq \left\| W^k - \overline{W} \right\|_G^2, \ \forall k \geq 0.$$

Thus, the sequence $\{W^k\}$ has a unique cluster point and

$$\lim_{k \to \infty} W^k = \overline{W}.$$

This completes the proof. $\qquad\square$

Remark: Here we choose γ^k as such to optimize the function (27). Actually this optimal

stepsize will not be too small even the iterate is close to the solution as shown in the following.

$$2\left(1-\alpha\sum_{i=1}^{m}\beta_i\right)\left\langle X^k - \widetilde{X}^k, \left(1-\alpha\sum_{i=1}^{m}\beta_i\right)\left(X^k - \widetilde{X}^k\right) - \alpha\left(C_0\left(X^k\right) - C_0\left(\widetilde{X}^k\right)\right)\right\rangle$$

$$\geq \quad \left(1-\alpha\sum_{i=1}^{m}\beta_i\right)^2\left\|X^k - \widetilde{X}^k\right\|^2 + \alpha^2\left\|C_0\left(X^k\right) - C_0\left(\widetilde{X}^k\right)\right\|^2$$

$$-2\left(1-\alpha\sum_{i=1}^{m}\beta_i\right)\alpha\left\langle X^k - \widetilde{X}^k, C_0\left(X^k\right) - C_0\left(\widetilde{X}^k\right)\right\rangle$$

$$= \quad \left\|R\left(X^k, \widetilde{X}^k\right)\right\|^2$$

$$\Rightarrow \quad \max\left\{1, 2\left(1-\alpha\sum_{i=1}^{m}\beta_i\right)\right\}\left\langle\begin{pmatrix}X^k - \widetilde{X}^k \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k\end{pmatrix}, \begin{pmatrix}R\left(X^k, \widetilde{X}^k\right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k\end{pmatrix}\right\rangle_G \geq \left\|\begin{pmatrix}R\left(X^k, \widetilde{X}^k\right) \\ Y_i^k - \widetilde{Y}_i^k \\ \lambda_i^k - \widetilde{\lambda}_i^k\end{pmatrix}\right\|_G^2$$

$$\Rightarrow \quad \gamma_*^k \geq \min\left\{1, \frac{1}{2\left(1-\alpha\sum\limits_{i=1}^{m}\beta_i\right)}\right\}. \tag{31}$$

Note that $\nu \in (0,2)$ is a relaxation factor put in front of it.

4   Numerical test on the completion problem of low-rank matrix

In this section, we present preliminary numerical results for the proposed alternating direction method for solving CNLSDP problems. We should emphasize that our purpose here is not to conduct extensive computational tests but to demonstrate that the proposed algorithm can be potentially efficient. The algorithm may be taken as prototypes of those sophisticated and tailor-made algorithms for solving different classes of problems.

In many fields of engineering and science, a low-rank matrix needs to be completed from small portion of entries observed. A good example is the well known Netflix problem [1]. This large online DVD renting company needs to provide recommendations to users based on their submitted ratings on some films. That means one would like to infer their preference for unrated items. This problem seems to be very hard in that we should fill in the missing entries of the matrix from only small samples. However, the matrix of all user-ratings to recover has low rank because there are only a few factors to explain an individual's preference for films. Then it can be modeled as follows.

$$\begin{aligned}\min \quad &\text{rank}(X) \\ \text{s.t.} \quad &X_{ij} = M_{ij}, \ (i,j) \in \Omega \\ &X \in \Re^{m\times n},\end{aligned} \tag{32}$$

where M is the unknown matrix and Ω is a set of pairs of indices for known entries.

To generalize, the affine rank minimization problem is introduced.

$$\begin{aligned}
\min \quad & \text{rank}(X) \qquad\qquad\qquad\qquad\qquad (33)\\
\text{s.t.} \quad & \mathcal{A}(X) = b\\
& X \in \Re^{m \times n},
\end{aligned}$$

where $\mathcal{A} : \Re^{m \times n} \to \Re^p$ is a linear operator and $b \in \Re^p$. This slight generalization appears useful in many areas such as machine learning [2], control [10, 13], and Euclidean embedding [41].

Notice that the affine rank minimization problem (33) is an NP-hard nonconvex optimization problem. A convex relaxation of (33) is given in [12] as follows.

$$\begin{aligned}
\min \quad & \|X\|_* \qquad\qquad\qquad\qquad\qquad\qquad (34)\\
\text{s.t.} \quad & \mathcal{A}(X) = b\\
& X \in \Re^{m \times n},
\end{aligned}$$

where $\|X\|_*$ is the nuclear norm of X. The nuclear norm of X is defined as

$$\|X\|_* = \sum_{i=1}^{q} \sigma_i(X),$$

where $q = \min\{m, n\}$ and $\sigma_i(X)$, $i = 1, \ldots, q$, are the singular values of X. Actually the nuclear norm is the best convex approximation of the rank function over the unit ball of matrices. Candes and Recht [5] proved that a random low-rank matrix can be recovered exactly with high probability from a rather small portion of entries by solving (34).

The problem (34) can be reformulated as an SDP problem,

$$\begin{aligned}
\min \quad & \frac{1}{2}\left(\langle W_1, I_m\rangle + \langle W_2, I_n\rangle\right) \qquad\qquad (35)\\
\text{s.t.} \quad & \mathcal{A}(X) = b\\
& \begin{pmatrix} W_1 & X \\ X^T & W_2 \end{pmatrix} \succeq 0.
\end{aligned}$$

In [38] SDPT3, one of the most advanced SDP solvers based on interior point methods, has been used to solve (35). However, the computational cost grows very fast as m and n increase.

The first-order methods may therefore provide a promising alternative to the interior point method due to their low sensitivity to problem sizes. Ma et al. [30] proposed a Bregman iterative algorithm for solving (34). Cai et al. [4] proposed a singular value thresholding algorithm for solving the Tikhonov regularized version of (34). More recently, Yang and Yuan suggested an inexact alternation direction method [44]. Toh and Yun [39] made an extensive computational experiment for the nuclear norm regularized least squares problem by using a proximal point gradient algorithm, which is similar in spirit to the alternation direction methods, and reported very promising computational results.

In applying our proposed alternating direction method to Problem (34), we substitute Step (6) by an equivalent minimization problem and apply the following lemma. For the proof, see Theorem 2.1 of [4] or Theorem 3 of [30].

Lemma 4.1. *The solution of the minimization problem*

$$\min_{X \in \Re^{m \times n}} \quad \frac{\tau}{2}\|X - G\|^2 + \mu\|X\|_*$$

for $\tau, \mu > 0$ is given in a closed form by

$$S_\tau(G) = U \, Diag\left(\left(\sigma - \frac{\mu}{\tau}\right)_+\right) V^T, \tag{36}$$

where $G = U\Sigma V^T$ and $\Sigma = Diag(\sigma)$ are from the singular value decomposition (SVD) of G.

Thus at each iteration of our proposed alternating direction method, there is actually an analytic solution and the main computational cost lies on computing the singular value decomposition (SVD) of a matrix. Furthermore, it suffices to know those singular values that are greater than the parameter $\frac{1}{\beta}$ and their corresponding singular vectors. If this parameter is large, the number of singular values to be evaluated is small. This motivates us to choose small β to make the decomposition of a large-scale matrix possible.

The random matrix completion problems considered in our numerical experiments are as follows.

Example. For each (n, r, p) triple, where n (we set $m = n$) is the matrix dimension, r is the predetermined rank, and p is the number of entries to sample, we generate $M = M_L M_R^T$ as in [39], where M_L and M_R are $n \times r$ matrices with i.i.d. standard Gaussian entries. Then a subset Ω of p elements uniformly at random from $\{(i, j) : i = 1, \ldots, n, j = 1, \ldots, n\}$ is selected. Therefore, the linear map $\mathcal{A}$ is given by

$$\mathcal{A}(X) = X_\Omega,$$

where $X_\Omega \in \Re^p$ are obtained from X by selecting those elements whose indices are in Ω. We take $\beta = 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.5, 1$ in order to observe the effect of parameter for $n/r = 100/10$. Then using $\beta = 0.1$, we test for $n/r = 200/10, 200/20, 500/10, 500/20, 500/50$, respectively, in order to observe the effect of problem size.

We choose the initial iterate to be $X^0 = Y^0 = \text{rand}(n)$ and $\lambda^0 = 0$. The stopping criterion we use is:

$$\frac{\left\|X^k - X^{k-1}\right\|_F}{\max\left\{\|X^k\|_F, 1\right\}} < 1\text{e-}4.$$

The accuracy of the computed solution X_{sol} by our algorithm can be measured by the relative error defined as follows:

$$\text{error} \equiv \frac{\left\|X_{\text{sol}} - M\right\|_F}{\|M\|_F},$$

where M is the original matrix.

For each case, we randomly generate 5 problems and report the average results of the alternating direction method in Tables 1 and 2. The columns corresponding to "ave. iter", "ave. #sv", and "ave. error" give the average number of iterations, the average number of nonzero singular values of the computed solution matrix, and the average relative error, respectively. As indicated in [4], an $n \times n$ matrix of rank r has $d_r \equiv r(2n - r)$ degrees of freedom. The ratio p/d_r is also shown in the tables, which indicates a degree of hardship in solving the problem.

The code was written in MATLAB (version 6.5) and the computations were performed on a 1.86 GHz Intel Core 2 PC with 3GB of RAM. In order to free ourselves from the distraction of having to consider the storage of too large matrices in MATLAB, we only use examples with moderate dimensions. Furthermore, we compute the full SVD at each iteration. From Table

Table 1: Numerical results with different β

Example	Unknown M			ADM		
$\beta=$	n/r	p	p/d_r	ave. iter	ave. #sv	ave. error
0.01	100/10	5666	3	135	19	1.4e-02
0.02	100/10	5666	3	83	18	5.6e-03
0.05	100/10	5666	3	53	13	5.3e-03
0.08	100/10	5666	3	63	11	7.0e-04
0.1	100/10	5666	3	71	10	3.5e-04
0.2	100/10	5666	3	106	10	1.2e-03
0.5	100/10	5666	3	202	11	3.7e-03
1	100/10	5666	3	351	12	8.2e-03

Table 2: Numerical results with $\beta = 0.1$

Example	Unknown M			ADM		
$\beta=$	n/r	p	p/d_r	ave. iter	ave. #sv	ave. error
0.1	200/10	15665	4	95	10	3.7e-04
0.1	200/20	22800	3	99	20	3.5e-04
0.1	500/10	49471	5	158	10	4.3e-04
0.1	500/20	78400	4	146	20	3.8e-04
0.1	500/50	142500	3	152	50	4.1e-04

1, it seems that $\beta = 0.1$ is a suitable parameter. Then using this β, the numerical results reported in Table 2 are competitive with those obtained by using the fixed point continuation algorithm and the accelerated proximal gradient algorithm in [39], which are proposed to solve easier unconstrained counterpart instead.

5 Conclusions

We propose a prediction-correction alternating direction method for solving convex nonlinear semidefinite optimization problems. The advantage of the proposed method is that it does not require to solve sub-variational inequality problems on semidefinite cone; instead, in each iteration, it requires only projections onto semidefinite cone plus m projections on convex sets. The convergence of the method is analyzed and it is shown that if the problem has an optimal solution at all, the method will produce a sequence that converge to a solution.

A numerical example of computing the low-rank completion of randomly generated matrix is presented. Our algorithm generates reasonably accurate solutions in a reasonable number of iterations in the experiment, showing that the proposed alternating direction method is promising in solving medium-sized convex nonlinear semidefinite optimization problems.

References

[1] ACM SICKDD AND NETFLIX, Proceedings of KDD Cup and Workshop, 2007. Avaiable online at http://www.cs.uic.edu/liub/KDD-cup-2007/proceedings.html.

[2] Y. Amit, M. Fink, M. Srebro, and S. Ullman, Uncovering shared structures in multiclass classification. *Proceedings of the Twenty-fourth International Conference on Machine Learning*, 2007.

[3] S. Burer and D. Vandenbussche, Solving lift-and-project relaxations of binary integer programs. *SIAM J. Optim.* 16, 726-750 (2006).

[4] J.F. Cai, E.J. Candes, and Z. Shen, A singular value thresholding algorithm for matrix completion. *Working Paper*, National University of Singapore, September 2008.

[5] E.J. Candes and B. Recht, Exact matrix completion via convex optimization. *Found. Comput. Math.* 9, 717-772 (2009).

[6] R. Correa and C.H. Ramirez, A global algorithm for nonlinear semidefinite programming. *SIAM J. Optim.* 15, 303-318 (2004).

[7] J. Douglas and H. Rachford, On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Socie.* 82, 421-439 (1956).

[8] J. Eckstein, Splitting methods for monotone operators with applications to parallel optimization. *PhD Thesis*, Massachusetts Institute of Technology, 1989.

[9] J. Eckstein, Parallel alternating direction multiplier decomposition of convex programs. *J. Optim. Theo. Appl.* 80, 39-62 (1994).

[10] L. El Ghaoui and P. Gahinet, Rank minimization under LMI constraints: A framework for output feedback problems. *Proceedings of the European Control Conference*, 1993.

[11] B. Fares, D. Noll, and P. Apkarian, Robust control via sequential semidefinite programming. *SIAM J. Contr. and Optim.* 40, 1791-1820 (2002).

[12] M. Fazel, Matrix rank minimization with applications. *PhD Thesis*, Stanford University, 2002.

[13] M. Fazel, H. Hindi, and S. Boyd, A rank minimization heuristic with application to minimum order system approximation. *Proceedings of the American Control Conference*, 2001.

[14] M. Fukushima, Application of the alternating directions method of multipliers to separable convex programming problems. *Compu. Optim. Appl.* 2, 93-111 (1992).

[15] D. Gabay, Applications of the method of multipliers to variational inequalities. In: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problem*, pp. 299-331. North-Holland, Amsterdam, The Netherlands, 1983.

[16] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Compu. Math. Appl.* 2, 17-40 (1976).

[17] M.S. Gowda, Y. Song, and G. Ravindran, On some interconnections between strict monotonicity, globally uniquely solvable, and P-properties in semidefinite linear complementarity problems. *Linear Algebra Appl.* 370, 355368 (2003).

[18] M.S. Gowda and R. Sznajder, Automorphism invariance of P- and GUS-properties of linear transformations on Euclidean Jordan algebras. *Math. Oper. Res.* 31, 109-123 (2006).

[19] B.S. He, L.Z. Liao, D.R. Han, and H. Yang, A new inexact alternating directions method for monotone variational inequalities. *Math. Program.* 92, 103-118 (2002).

[20] B.S. He, L.Z. Liao, and M.J. Qian, Alternating projection based prediction-correction methods for structured variational inequalities. *J. Compu. Math.* 24, 683-710 (2006).

[21] F. Jarre, An interior method for nonconvex semidefinite programs. *Optim. and Eng.* 1, 347-372 (2000).

[22] C. Kanzow, C. Nagel, H. Kato, and M. Fukushima, Successive linearization methods for nonlinear semidefinite programs. *Compu. Optim. Appl.* 31, 251-273 (2005).

[23] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications.* Academic Press, New York, 1980.

[24] M. Kocvara and M. Stingl, PENNON: a code for convex nonlinear and semidefinite programming. *Optim. Meth. Soft.* 18, 317-333 (2003).

[25] L. Kong, J. Sun, and N. Xiu, A regularized smoothing Newton method for symmetric cone complementarity problems. *SIAM J. Optim.* 19, 1028-1047 (2008).

[26] S. Kontogiorgis, Alternating directions methods for the parallel solution of large-scale block-structured optimization problems. *PhD Thesis*, University of Wisconsin-Madison, 1994.

[27] S. Kontogiorgis, R.D. Leone, and R.R. Meyer, Alternating direction splittings for block angular parallel optimization. *J. Optim. Theo. Appl.* 90, 1-29 (1996).

[28] F. Leibfitz and M.E. Mostafa, An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems. *SIAM J. Optim.* 12, 1048-1074 (2002).

[29] P. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* 16, 964-979 (1979).

[30] S. Ma, D. Goldfarb, and L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization. *Working Paper*, Columbia University, October 2008.

[31] D. Noll and P. Apkarian, Spectral bundle methods for non-convex maximum eigenvalue functions: first-order methods. *Math. Program.* 104, 701-727 (2005).

[32] D. Noll and P. Apkarian, Spectral bundle methods for non-convex maximum eigenvalue functions: second-order methods. *Math. Program.* 104, 729-747 (2005).

[33] J. Povh, F. Rendl, and A. Wiegele, A boundary point method to solve semidefinite programs. *Computing* 78, 277-286 (2006).

[34] R.T. Rockafellar and R.J.B. Wets, Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.* 16, 119-147 (1991).

[35] D. Sun, J. Sun, and L. Zhang, Rates of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math. Program.* 114, 349-391 (2008).

[36] J. Sun, D. Sun, and L. Qi, A squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems. *SIAM J. Optim.* 14, 783-806 (2004).

[37] J. Sun, K.C. Toh, and G. Zhao, An analytic cutting plane method for semidefinite feasibility problems. *Math. Oper. Res.* 27, 332-346 (2002).

[38] K.C. Toh, M.J. Todd, and R.H. Tutuncu, SDPT3 - a MATLAB software package for semidefinite-quadratic-liear programming, http://www.math.nus.edu.sg/̃mattohkc/sdpts.html.

[39] K.C. Toh and S. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. To appear in *Pacific J. Optim.*

[40] K.C. Toh, G. Zhao, and J. Sun, A multiple-cut analytic cutting plane method for semidefinite feasibility problems. *SIAM J. Optim* 12, 1126-1146 (2002).

[41] M.W. Trosset, Distance matrix completion by numerical optimization. *Comput. Optim. Appl.* 17, 11-22, (2000).

[42] Z. Wen, D. Goldfarb, and W. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming. *Optim. Online*, Aug. 2009.

[43] H. Yamashita, H. Yabe, and K. Harada, A primal-dual interior point method for nonlinear semidefinite programming. *Technical Report*, Tokyo University of Science, 2007.

[44] J.F. Yang and X.M. Yuan, An inexact alternating direction method for trace norm regularized least squares problem, under revision of *Mathematics of Computation*, 2011.